УДК 004.934.5

Т.В. Людовик
Международный научно-учебный центр информационных технологий
и систем НАН Украины и МОН Украины, г. Киев, Украина
tetyana.lyudovyk@gmail.com

Автоматический синтез нейтральной и выразительной речи

В интеллектуальных приложениях, использующих речевые технологии, синтезированная речь должна звучать естественно и выразительно. В статье описана разработанная технология синтеза речи, обеспечивающая озвучивание произвольных орфографических текстов на украинском языке в нейтральном и выразительном стилях с сохранением индивидуальных особенностей голоса и произношения. Основное внимание уделено просодической модели интонирования, используемой для синтеза речи с нейтральной и выразительной интонацией.

Введение

Технология автоматического синтеза речи по тексту (TTS, Text-to-Speech) позволяет передавать голосовую информацию от компьютера к человеку, преобразовывая произвольный орфографический текст в звучащую речь.

К современным системам синтеза речи (TTS-системам) предъявляются требования разборчивости и естественности (натуральности) звучания. Разборчивость подразумевает правильное распознавание человеком всех слов синтезированной речи. Большинство современных TTS-систем демонстрируют хорошую разборчивость, приближающуюся к разборчивости естественной речи. В то же время практика показывает, что разборчивая, но неестественно звучащая речь не удовлетворяет требованиям пользователей. Естественность синтезированной речи оценивается по тому, насколько она похожа на речь живого человека, насколько она выразительна и насколько в ней отражены индивидуальные особенности голоса и произношения.

Индивидуальность естественной речи проявляется как в физиологических особенностях голоса, так и в приобретенных привычках произношения.

Под выразительностью речи понимается выражение отношения читающего текст к содержанию этого текста и к аудитории. Как правило, стиль и смысл текста диктуют выбор стиля речи. В выразительной речи подчеркнуты отдельные слова, выделены паузами определенные участки текста и т.д. Между нейтральной и выразительной речью нет четких границ, речь может идти о различной степени выразительности.

Эмоциональность речи связана с состоянием говорящего. Характеризуя в основном спонтанную речь, при чтении вслух эмоциональность, как правило, имитируется, добавляя выразительности.

В большинстве приложений, не требующих интерактивного взаимодействия человека с компьютером, уместна нейтрально звучащая синтезированная речь (озвучивание новостной, навигационной, деловой информации). Более выразительная речь необходима для озвучивания художественной литературы и в ситуациях живого общения человека с компьютером: в интерактивных образовательных программах, играх, в разговоре с роботами-собеседниками. В этих и иных интерактивных приложениях,

таких как автоматизация работы call-центров, внимание в основном уделяется проблеме распознавания речи и эмоций. Однако адекватный синтез речи дружелюбным, уместно жизнерадостным или сочувственным голосом также играет большую роль. Выразительная речь компьютера стимулирует пользователя поддерживать с ним взаимодействие.

Крупные фирмы, работающие в области синтеза речи, активно разрабатывают и рекламируют «живые», «выразительные», «приятно звучащие» голоса. Фирма **Acapela** разработала пользующийся успехом многоязычный сервис «говорящие открытки», а также несколько голосовых приложений для iPhone, позволяющих читать новости, переводить тексты, изучать иностранный язык. Фирма **Nuance** предлагает фирмам-заказчикам создавать корпоративные, узнаваемые, ассоциируемые с фирмой-заказчицей естественно звучащие голоса. **Loquendo** также предлагает выразительные, наделенные чертами конкретных людей голоса. На рынок TTS-систем выходит также российский разработчик **Центр речевых технологий** (ЦРТ).

Синтез нейтральной и выразительной речи методом Unit Selection

Различие между нейтральным и выразительным стилями чтения проявляется в основном на уровне просодики – интонации, ритмики, паузации, темпа произнесения текста и его отдельных частей, а также степени ударения отдельных слов и слогов. В речевом сигнале просодическим характеристикам соответствуют акустические характеристики: частота основного тона (fundamental frequency, F0), длительность и интенсивность (амплитуда).

Исследования, посвященные синтезу выразительной речи, можно разделить на несколько направлений: синтез речи в определенном стиле [1], синтез эмоциональной речи [2], синтез эмфатической речи (с логическим выделением отдельных слов) [3], синтез характерной речи (с признаками индивидуальных особенностей характера) [4]. Объединяет эти исследования то, что в обязательном порядке моделируются интонация (контур F0), а также длительность звуков и пауз. Тестирование синтезированной речи [5] показало, что спектральные характеристики, в отличие от просодических, менее важны.

В настоящее время наиболее распространенным методом синтеза речи является метод Unit Selection [6], основанный на генерации речевого сигнала путем конкатенации естественных речевых отрезков, выбираемых из речевой базы данных. В речевой базе данных, содержащей отрезки речи конкретного диктора и имеющей большой объем (5 – 15 часов акустических записей), отражаются как особенности голоса этого диктора, так и используемый им стиль чтения. Как правило, используются речевые отрезки, соответствующие отдельным звукам или дифонам (участкам речи от середины одного звука до середины следующего). Большое количество элементов (units) базы данных, различающихся спектральными и просодическими характеристиками, позволяет синтезировать речь с высокой степенью естественности. Чем больше объем речевой базы данных, тем с большей вероятностью в ней будут найдены необходимые для синтеза речевые отрезки и меньше придется модифицировать синтезированный сигнал, добиваясь необходимых значений длительности, F0 и плавных переходов от одного звука к другому. Известно, что любая модификация речевого сигнала отрицательно сказывается на качестве его звучания. Речь, синтезируемая методом Unit Selection, сохраняет индивидуальные особенности голоса и произношения диктора.

Важной составляющей метода Unit Selection является алгоритм выбора элементов из базы данных. Проблема состоит в том, что приходится решать, какие критерии выбора важнее: контекст, интонация, длительность и т.д. Поскольку сбалансированность критериев не достигнута, а выбор осуществляется автоматически, процесс синтеза речи иногда «выходит из-под контроля» [7], и синтезированная речь воспринимается как «неуравновешенная».

Теоретически существует три способа, не отказываясь от метода Unit Selection, обеспечивающего естественное звучание, синтезировать речь в различных стилях:

1) создать параллельно несколько речевых баз данных на материале одного голоса, но разных стилей;

2) аннотировать единую базу данных метками, соответствующими стилям, и учитывать эту информацию в процессе синтеза при выборе речевых элементов из единой базы данных;

3) разработать просодические модели, настраиваемые на голос и стиль, использовать их для порождения индивидуальных и стилевых значений просодических характеристик, учитываемых при выборе речевых элементов в процессе синтеза речи.

Более перспективными представляются второй и третий, наименее трудоемкий, подходы.

Цель исследования

Целью данной работы является разработка технологии синтеза речи, обеспечивающей озвучивание произвольных орфографических текстов на украинском языке в нейтральном и выразительном стилях с сохранением индивидуальных особенностей голоса и произношения. Основное внимание уделено разработке просодической модели интонирования, позволяющей синтезировать речь с нейтральной и выразительной интонацией.

Синтез украинской нейтральной и выразительной речи по тексту

В системе синтеза украинской речи [8] используется разработанный в МНУЦИТиС фонемно-трифонный метод синтеза речи в амплитудно-временной области, являющийся вариантом метода Unit Selection. Объединение метода синтеза с разработанными индивидуализированными просодическими моделями позволяет озвучивать тексты в соответствии с выбранными голосами и стилями чтения.

Разработанная система синтеза индивидуализированной украинской речи (рис. 1) состоит из следующих компонентов:

– речевых баз данных;
– лингвистического процессора;
– модуля выбора элементов из речевой базы данных;
– акустического процессора.

Речевые базы данных используются не только в процессе синтеза речи. Содержащаяся в их аннотациях информация служит для предварительной настройки моделей произношения диктора. В процессе синтеза речи настроенный лингвистический процессор генерирует фонемно-просодическую транскрипцию входного текста в виде последовательности фонем с вычисленными просодическими характеристиками длительности и интонационного контура.

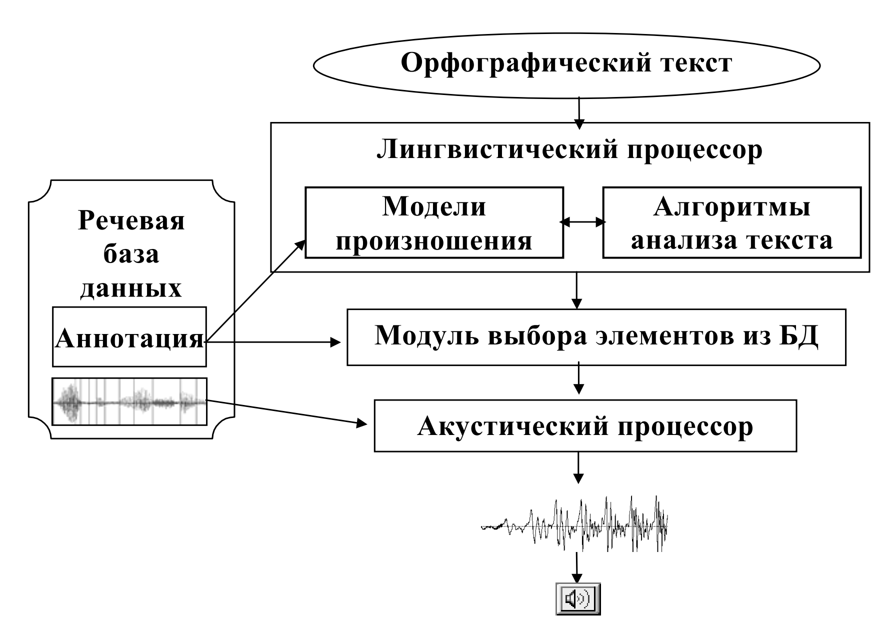


Рисунок 1 – Блок-схема системы синтеза украинской речи по тексту

Модуль выбора элементов из базы данных сравнивает фонемно-просодическую транскрипцию входного текста (то есть информацию о том, что и как должно синтезироваться) с аннотацией базы данных (то есть с информацией о том, какой речевой материал имеется в наличии). Модуль выбора оценивает и выбирает элементы речевой базы данных в соответствии с характеристиками, определенными при анализе текста.

Выбранные элементы конкатенируются акустическим процессором и озвучиваются акустической системой.

Речевые базы данных

Качество синтезированной речи зависит от объема и покрытия речевой базы данных (РБД), то есть от того, насколько полно в ней представлены звуковые, темпоральные и интонационные варианты речевых единиц.

Элементами РБД являются аллофоны (фонемы-трифоны), то есть фонемы в сегментном контексте, указывающем, какие фонемы находятся в речевом сигнале слева и справа от данной фонемы. Каждый элемент аннотирован идентификатором, именем, состоящим из трех частей (имя предыдущей, текущей и последующей фонемы), длительностью, значениями интенсивности, для гласных и звонких согласных также последовательностью длин периодов основного тона и количеством периодов.

На данный момент разработаны 7 РБД: 4 мужских голоса, 2 женских и 1 детский (табл. 1). Наименее естественно звучит речь, синтезируемая на основе РБД, состоящей из прочитанного диктором словаря объемом около 300 слов. Разборчивость такой речи, тем не менее, довольно высокая благодаря тому, что в словаре представлены все фонемы украинского языка в наиболее часто встречающихся контекстах. Наиболее естественно звучит речь, синтезируемая с использованием РБД, разработанной на основе текстов различных типов. РБД диктора НАТАЛКИ содержит речь различных стилей чтения, что позволяет синтезировать речь в дипазоне от наиболее нейтральной (прогноз погоды) до наиболее выразительной (диалоги).

Таблица 1 – Речевые базы данных, используемые в системе синтеза украинской речи

Диктор	Тип текста	Продолжительность акустической записи (минуты)	Объем речевой базы данных (количество аллофонов)
СВЯТОСЛАВ	Изолированные фразы	26 мин. 16 с.	11843
ДМИТРИЙ	Радионовости	8 мин. 57 с.	8433
МАРИНА	Изолированные слова и фразы	10 мин. 44 с.	3017
ОСТАП	Изолированные слова	4 мин. 42 с.	1424
ДАРИНКА	Изолированные слова и фразы	5 мин. 28 с.	2384
ПАНАС	Художественный текст	79 мин. 30 с.	41213
НАТАЛКА	Художественный текст, тексты SMS-сообщений, диалоги, радионовости, изолированные фразы, названия букв алфавита, даты, телефонные номера и т.п.	234 мин. 22 с.	70478

Лингвистический анализ

В процессе лингвистического анализа входного текста задействованы дикторонезависимые алгоритмы обработки текста и дикторозависимые модели произношения. Для настраивания моделей используются аннотации индивидуальных РБД. Учитывается, как диктор произносит те или иные звукосочетания (ассимиляция, редукция), где ставит ударения (например, «зокрЕма» или «зокремА»), как часто делает паузы внутри фраз. Наиболее важной, особенно для синтеза выразительной речи, является индивидуализация моделей длительности и интонации.

Модель длительности фонем

Вычисление длительности фонем осуществляется с помощью модели, параметрами которой являются: средняя длительность фонемы (по аннотации РБД), тип контекста, в котором она находится в синтезируемом высказывании, и набора коэффициентов длительности для данной фонемы, соответствующих типу контекста. В процессе синтеза речи тип контекста устанавливается с учетом коммуникативного типа синтагмы, наличия в синтагме логического ударения, позиции фонемы по отношению к началу/концу синтагмы, типа слога (открытый, закрытый) и сегментного типа непосредственного левого и правого окружения (ударная/безударная гласная, согласная фонемы). Для вычисления длительности фонемы ее средняя длительность умножается на коэффициент, соответствующий типу контекста.

Модель длительности фонем индивидуализируется автоматически.

Модель интонирования

Модель интонирования используется для вычисления интонационных контуров – последовательностей значений F0 на протяжении текста. Модель основана на том, что главной интонационной единицей речи считается синтагма – часть фразы, имеющая выраженный интонационный контур. Синтагма состоит из одной или нескольких акцентных групп. Акцентная группа (акцентная единица) – это одно или несколько слов, объединенных общим ударением. Разработанная модель интонирования близка к модели интонационных портретов акцентных единиц, предложенной Б.М. Лобановым [9].

Параметрами интонационной модели являются:

– коммуникативный тип синтагмы, определяемый в настоящее время по знаку пунктуации и некоторым лексико-грамматическими показателям (вопросительные слова, союзы и т.д.);
– количество акцентных групп в синтагме;
– место ядерной (главной) акцентной группы в синтагме;
– наборы целевых значений F0 для каждой акцентной группы.

В любой синтагме обязательно присутствует ядерная акцентная группа (АГ), несущая главное (синтагматическое) ударение. В общем случае, если в синтагме две АГ, то первая из них является начальной, а вторая – ядерной. Если акцентных групп три или больше, то первая из них является начальной, со второй по предпоследнюю включительно – предъядерной, последняя – ядерной. Наличие логического ударения в синтагме может сделать ядерной любую АГ, в этом случае все АГ, следующие за ядерной, считаются заядерными.

Каждый коммуникативный тип синтагмы имеет свой интонационный контур, состоящий из интонационных контуров входящих в нее АГ. Каждая АГ синтагмы состоит из ядра – ударной гласной, предъядра – всех фонем АГ, находящихся перед ударной гласной, и заядра – всех фонем АГ, находящихся после ударной гласной. Главное предположение модели интонирования состоит в том, что топологические свойства просодических характеристик не изменяются (или изменяются незначительно) с изменениями фонетического контекста и числа фонем в предъ- и заядре АГ [9].

Контур АГ задается последовательностью 10 значений F0. Контур синтагмы задается 10n значениями F0, где n – количество АГ в синтагме. Интонационные контуры акцентных групп синтагмы «накладываются» на их фонемные транскрипции, каждое из 10 целевых значений F0 приписывается соответствующим целевым точкам АГ. Первые два из 10 целевых значений F0 задают движение F0 на предъядре АГ; значения F0 с 3 по 8 задают изменение F0 на ядре (ударной гласной); последние два значения F0 описывают движение F0 на заядре АГ.

На первом этапе индивидуализации проводится классификация синтагм, произнесенных диктором и хранящихся в РБД, по их коммуникативному типу (КТ). Минимальный набор учитываемых коммуникативных типов: завершенность («Знову залягла мертва тиша.»); завершенность с логическим ударением («Саме вона працює.»); незавершенность («Аліса відчула,»); общий вопрос («Це місце вільне?»); специальный вопрос («Як твоє самопочуття?»); восклицание («Який божевільний день!»); перечисление («Бажаю здоров’я, щастя, добра.»); противопоставление («Це не гра, а реальність.»). Модель позволяет увеличивать количество коммуникативных типов за счет детализации коммуникативного смысла (например, собственно вопрос, переспрос и т.д.).

Синтагмы выделяются в автоматизированном режиме с последующим аудитивным и визуальным контролем. Далее проводится классификация синтагм в рамках

одного КТ по количеству АГ и месту ядерной АГ. Каждый подтип имеет название, состоящее из трех частей: кода КТ, количества АГ и места ядерной АГ в синтагме, например, Z-3-2 (завершенный тип, три АГ в синтагме, вторая АГ является ядерной).

На следующем этапе индивидуализации модели интонирования проводится стилизация (упрощение) интонационных контуров синтагм с помощью программных средств обработки интонационной части аннотаций РБД. Стилизация осуществляется в соответствии с описанной моделью интонирования: определяются значения F0 в целевых точках акцентных групп синтагмы: для каждой АГ определяются два значения на предъядре, шесть значений на ядре и два значения на заядре. Результаты стилизации удобно представлять в виде таблиц и графиков, где на оси абсцисс отложены целевые точки АГ (A(1,1), A(1,2), … A(n,9), A(n,10), где n – количество АГ в синтагме), а на оси ординат – целевые значения F0 в Гц. Стилизация дает возможность сравнивать интонационные контуры синтагм различного сегментного (фонемного) состава, а также интонационные контуры разных дикторов. На рис. 2 приведены стилизованные интонационные контуры синтагмы завершенного типа «Скоро всю землю розгородимо.» в произнесении четырех дикторов.

Особенности произношения проявляются в том, что дикторы по-разному интерпретируют одну и ту же фразу: РУСЛАН и АЛЕКСАНДР произносят нейтрально (движение F0 восходящее – восходящее – нисходящее у АЛЕКСАНДРА и восходящее – ровное – нисходящее у РУСЛАНА), а СВЯТОСЛАВ и МАКСИМ выделяют слово «землю» (восходяще-нисходящее движение F0 на второй АГ), придавая выразительный, экспрессивный характер всей синтагме. Произношение РУСЛАНА отличается монотонностью. Интонационные различия произношения дикторов свидетельствуют о том, что информации о коммуникативной значимости отдельных частей синтагмы в тексте содержится недостаточно и она неоднозначна.

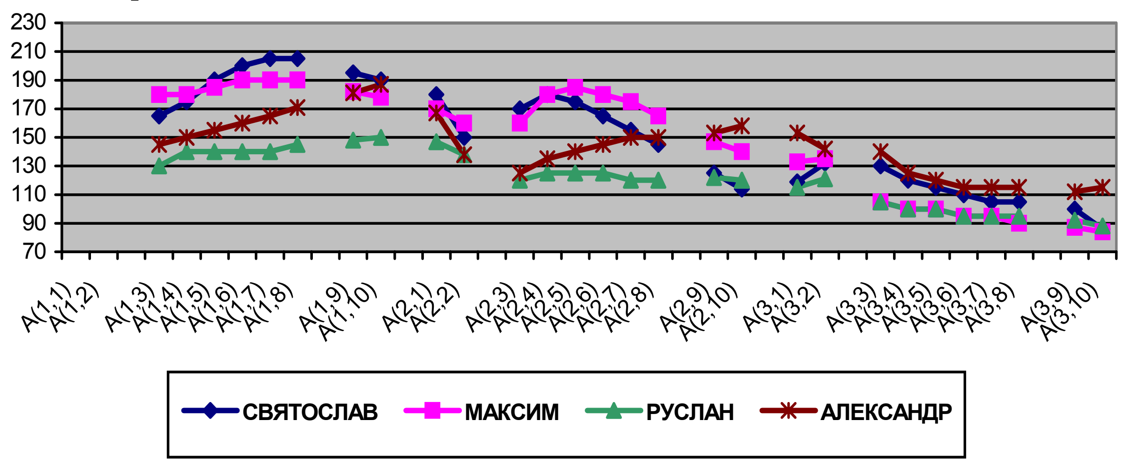


Рисунок 2 – Стилизованные интонационные контуры синтагмы завершенного типа «Скоро всю землю розгородимо.» в произнесении четырех дикторов

Выразительность речи связана не только с переносом ядерной АГ с конца синтагмы. Как правило, выразительный характер синтагмы проявляется в более широком диапазоне F0 всей синтагмы или ее отдельных АГ по сравнению с нейтральной речью. На рис. 3 приведены интонационные контуры нейтрально прочитанной диктором НАТАЛКОЙ синтагмы завершенного типа Z-3-3 «Без істотних опадів» и выразительно прочитанной синтагмы аналогичного подтипа «Це моя сестра Іра» (слова «Це» и «моя» составляют одну АГ). В обоих случаях ядерной остается последняя АГ синтагмы.

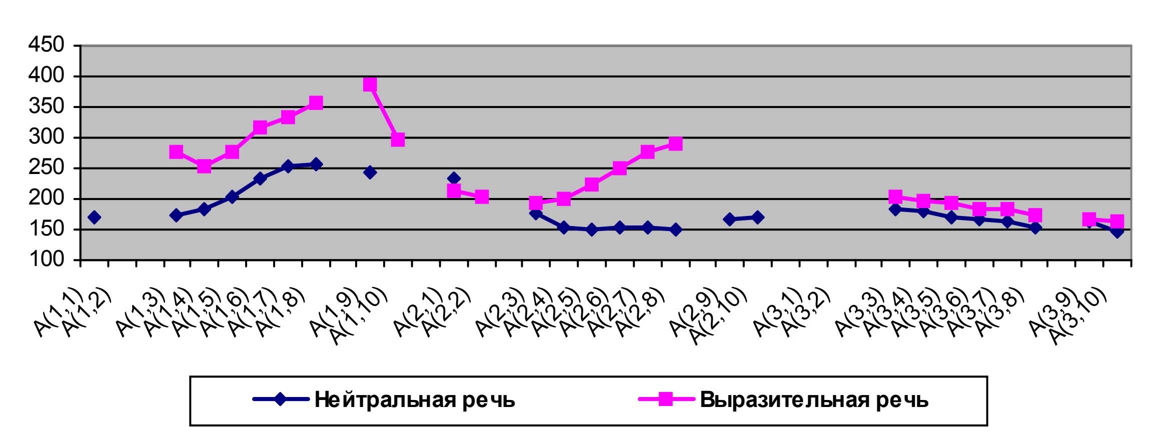


Рисунок 3 – Интонационные контуры нейтральной и выразительной речи диктора НАТАЛКИ

Рис. 4 свидетельствует о вариативности речи одного и того же диктора. Интонационные контуры одного подтипа в произнесении одного диктора различаются уровнем и диапазоном и направлением движения F0 на начальной и предъядерной АГ, однако общим для всех приведенных контуров является восходящее движение F0 на ядерной АГ, что и позволяет воспринимать интонацию синтагмы как незавершенную.

После распределения всех синтагм РБД диктора по коммуникативным подтипам, классификации синтагм в соответствии с нейтральным/выразительным стилем и стилизации интонационных контуров синтагм выводится интонационная модель данного диктора. Это происходит либо путем усреднения значений F0 в целевых точках синтагмы, либо в качестве модельного контура синтагмы эмпирическим путем выбирается один из стилизованных контуров.

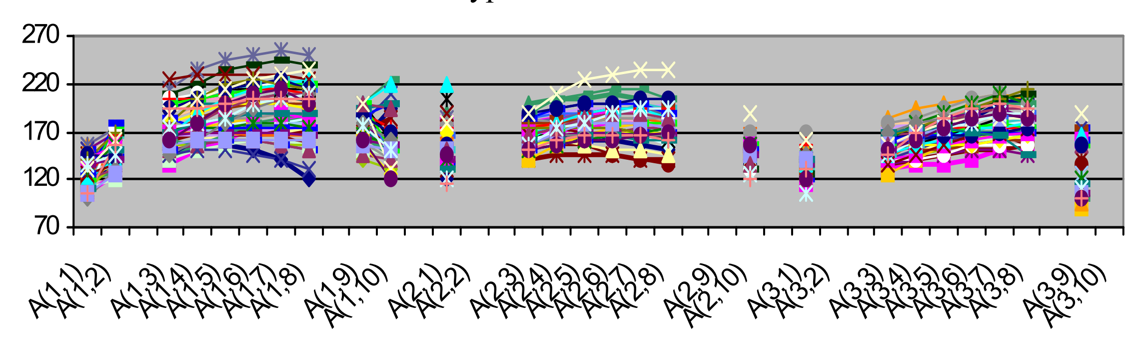


Рисунок 4 – Стилизованные интонационные контуры синтагм подтипа незавершенности N-3-3 в произнесении диктора СВЯТОСЛАВА

Множество модельных интонационных контуров диктора, репрезентирующих все коммуникативные подтипы и стили (нейтральный и выразительный), представляет собой индивидуализированную модель интонации этого диктора.

На рис. 5 приведены модельные интонационные контуры синтагмы нейтральной завершенности подтипа Z-3-3 в произнесении двух дикторов. Приведенные контуры нормированы: ось ординат соответствует не абсолютным, а нормированным значениям F0. Нормированное значение «0» соответствует 150 Гц для диктора НАТАЛКИ и 80 Гц для диктора ПАНАСА. Максимальное значение «10» соответствует 375 Гц для НАТАЛКИ и 180 Гц для ПАНАСА. Более широкий диапазон интонационного контура НАТАЛКИ свидетельствует о большей выразительности этого голоса.

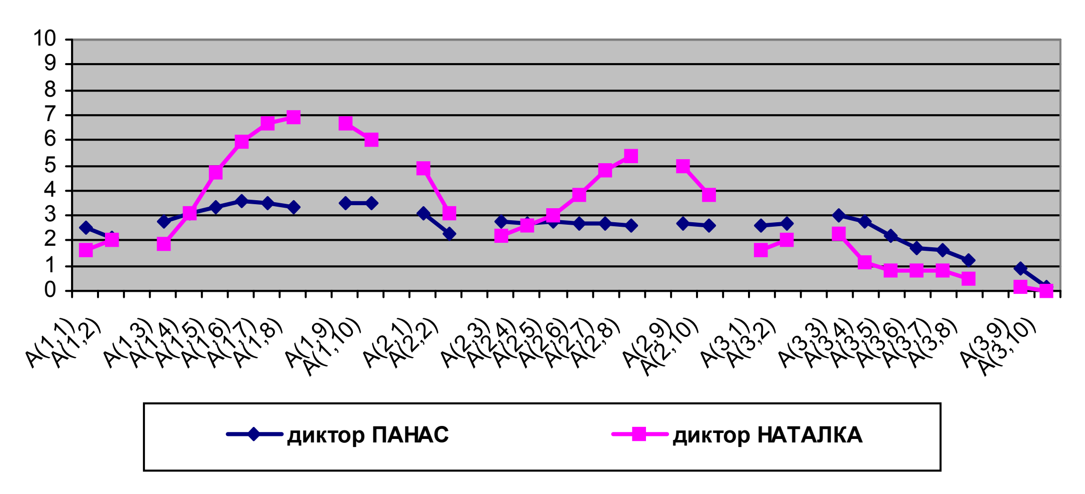


Рисунок 5 – Модельные интонационные контуры завершенности подтипа Z-3-3 дикторов ПАНАСА и НАТАЛКИ

В процессе синтеза речи индивидуализированные модели интонации используются лингвистическим процессором для интонирования входного орфографического текста. Решение о том, контур какого коммуникативного подтипа выбирать для озвучивания синтагмы в процессе синтеза речи, принимается с учетом знака пунктуации и количества акцентно-ударных гласных в синтагме после ее вычленения в тексте. По умолчанию ядерной считается последняя акцентная группа синтагмы. Однако пользователю системы синтеза речи предоставляется возможность при вводе текста обозначать логическое выделение особым знаком логического ударения.

Результаты тестирования синтезированной речи

Было проведено формальное тестирование образцов синтезированной речи с целью определить, какой из голосов, ПАНАС или НАТАЛКА, звучит естественнее. В частности, обоими голосами были озвучены отрывки из художественного, публицистического и научного текстов, а также тексты 33 SMS-сообщений. В тестировании методом средней оценки (MOS, mean opinion score) [10] участвовали сотрудники МНУЦИТиС (5 человек), а также преподаватели и студенты специальности «Украинский язык и литература» (33 человека). Всем аудиторам предлагалось прослушать образцы синтезированной речи, предъявляемые в случайном порядке, и выставить каждому образцу оценку от 0 (очень плохо) до 5 (очень хорошо). Средние оценки естественности речи, синтезированной двумя голосами, приведены на рис. 6.

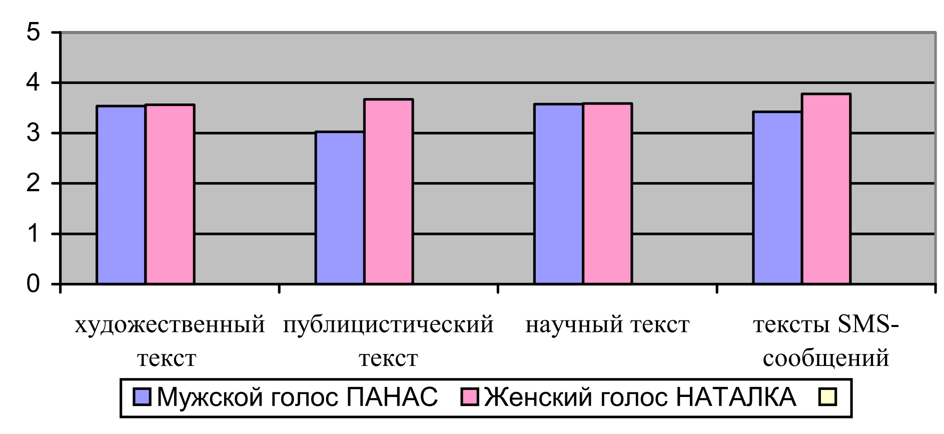


Рисунок 6 – Результаты тестирования естественности синтезированной речи

Обе группы аудиторов отдали незначительное предпочтение женскому голосу НАТАЛКА. Наименьшее различие (0,01 балла) наблюдается в оценке научного текста. Это может быть объяснено тем, что научный стиль не требует выразительного чтения, присущего НАТАЛКЕ, а строгость этого стиля в большей степени ассоциируется с мужским голосом. Незначительное различие (0,03 балла) в оценке художественного текста может быть объяснено тем, что желательная выразительность женского голоса нивелируется недостаточной степенью контроля над просодикой при выборе единиц конкатенации, в результате чего нарушается восприятие тонких смысловых связей, присущих художественным текстам. Предпочтение женского выразительного, хотя и более «неуравновешенного» голоса, мужскому нейтральному, хотя и более стабильному, заметнее для публицистического текста (0,7 балла) и озвучивания SMS-сообщений (0,4 балла).

Выводы

Синтезированная речь может считаться качественной, если она не только разборчива, но и звучит естественно. Естественность синтезированной речи связана с ее выразительностью и отображением индивидуальных особенностей произношения.

Разработанная технология синтеза речи решает проблему индивидуализации и стиля чтения (нейтральный/выразительный) благодаря предварительному настраиванию системы синтеза речи.

Тестирование системы синтеза украинской речи независимыми экспертами дало положительные результаты.

Литература

1. The IBM expressive speech synthesis system / [Hamza W., Eide E., Bakis R., Picheny M., Pitrelli J.] // Proceedings of the International Conference INTERSPEECH-2004. – P. 2577-2580.
2. Towards emotional speech synthesis: a rule based approach / [Zovato E., Pacchiotti A., Quazza S., Sandri S.] // Proceedings of the 5th ISCA workshop on Speech Synthesis. – Pittsburgh, 2004. – P. 219-220.
3. Modelling prominence and emphasis improves unit-selection synthesis / [Strom V., Nenkova A., Clark R. и др.] // Proceedings of the International Conference INTERSPEECH-2007 (Antwerp, (Belgium), August 2007). – P. 1282-1285.
4. Modelling personality features by changing prosody in synthetic speech / [Trouvain J., Schmidt S., Schröder M. и др.] // Proceedings of the Conference on Speech Prosody. – Dresden (Germany), 2006. – P. 088.
5. Emphasized speech synthesis based on hidden Markov models / [Morizane K., Nakamura K., Toda T. и др.] // Proceedings of the Oriental COCOSDA International Conference on Speech Database and Assessments. – Beijing (China), Aug. 2009. – P. 76-81.
6. Hunt A. Unit selection in a concatenative speech synthesis system using a large speech database / Hunt A., Black A. // Proceedings of the International Conf. on Acoustics, Speech, and Signal Processing. – Atlanta (USA), 1996. – Vol. 1. – P. 373-376.
7. Perspective on the Next Challenges for TTS Research / [Schroeter J., Conkie A., Syrdal A. и др.] // Proceedings of the IEEE Workshop on Speech Synthesis. – 2002. – P. 211-214.
8. Lyudovyk T. Unit Selection Speech Synthesis Using Phonetic-Prosodic Description of Speech Databases / Lyudovyk T., Sazhok M. // Proceedings of the International Conference «Speech and Computer» (SPECOM'2004). – St.-Petersburg (Russia), 2004. – P. 594-599.
9. Лобанов Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск : Белорус. Наука, 2008. – 337 с.
10. Sityaev D. Comparison of the ITU-t p.85 standard to other methods for the evaluation of text-to-speech systems / Sityaev D., Knill K., Burrows T. // Proceedings of the International Conference INTERSPEECH-2006. – P. 1077-1080.

Людовик Т.В.
Автоматичний синтез нейтрального та виразного мовлення
В інтелектуальних прикладних розробках, в яких використовуються мовленнєві технології, синтезоване мовлення повинне звучати натурально та виразно. У статті описується розроблена технологія синтезу мовлення, яка забезпечує озвучення довільних орфографічних текстів українською мовою у нейтральному та виразному стилях. Основна увага приділяється просодичній моделі інтонування, яка використовується для синтезу мовлення з нейтральною та виразною інтонацією.

Статья поступила в редакцию 22.01.2010.